

# HOOFDSTUK 4 DE WAARNEMING VAN SPRAAK

*Vincent J. van Heuven*

*Vakgroep Algemene Taalwetenschap/Fonetisch Laboratorium, Rijksuniversiteit Leiden, Postbus 9515  
2300 RA Leiden*

## SAMENVATTING

Dit hoofdstuk behandelt de bouw en werking van het menselijk oor met speciale aandacht voor de waarneming van eigenschappen van spraakgeluiden. Tevens wordt de gevoeligheid van het oor voor klankverschillen besproken. Tenslotte wordt ingegaan op de mogelijke unieke status van spraakwaarneming in de menselijke perceptie.

## 1 INLEIDING

De waarneming van spraak is een ingewikkeld proces waarvan we nog maar een fractie begrijpen. Spraakwaarneming is een bijzondere vorm van geluidswaarneming. De waarneming van spraak verloopt dan ook vooral via het gehoor. Overigens merken we hierbij meteen op dat spraak ook een visueel aspect heeft. Een luisteraar weet niet alleen hoe spraak moet klinken, hij weet ook welke bewegingen een spreker met zijn kaak en lippen maakt om een bepaalde klankreeks voort te brengen. Als het beeld van de spreker en het spraakgeluid niet met elkaar sporen, zoals bij een slechte (na)synchronisatie op de televisie, dan merken we dat meteen, en storen ons daaraan. Er zijn aanwijzingen dat in zulke gevallen de verstaanbaarheid van de spraakboodschap te lijden heeft [13, 20, 30]. Toch zal het duidelijk zijn dat het geluidsaspect bij het spraakverstaan overweegt. Blinden verstaan spraak moeiteloos, doven houden ook na een training in liplezen veel problemen. In dit hoofdstuk beperken we de waarneming van spraak dan ook tot het geluidsaspect.

Om spraak te kunnen verstaan moeten we spraak kunnen horen. Horen doen we met onze oren. In het hoofdstuk van 'T HART is verteld dat (spraak)geluid als akoestisch verschijnsel niets anders is dan kleine, maar heel snel op elkaar volgende, verstoringen van de luchtdruk. Zulke luchtdrukverstoringen (of trillingen) worden in ons oor omgezet in *zenuwimpulsen*, die onze hersenen vertellen dat er geluid klinkt. In uitzonderlijke gevallen kunnen zulke zenuwimpulsen spontaan optreden zonder dat er akoestisch geluid aan te pas komt. Iedereen heeft wel eens een 'piep in zijn oor', ook zijn er gehoorstoornissen (bijvoorbeeld *tinnetus aures*) waarbij de patient constant geplaagd wordt door een luid gerinkel waar geen aanwijsbare akoestische oorzaak voor is. Wij zullen in dit hoofdstuk onder meer nagaan hoe akoestische trillingen in het oor worden omgezet in zenuwimpulsen, en welke eigenschappen van de geluidstrillingen in deze zenuwimpulsen bewaard blijven. Eigenschappen die niet op een of andere manier vertaald worden in zenuwimpulsen worden niet gehoord en kunnen geen invloed hebben op het proces van spraakverstaan. Omdat de materie al ingewikkeld genoeg is zal een aantal complicaties bij het spraakverstaan buiten beschouwing blijven.

Mensen weten feitloos wat spraakgeluid is en wat niet. Het is op dit ogenblik niet mogelijk precies te definiëren welke eigenschappen geluiden moeten bezitten om als menselijk spraakgeluid herkend te worden.

Menselijke luisteraars hebben vrijwel onmiddellijk in de gaten of zij in hun eigen taal of in een andere taal worden toegesproken. Het is niet bekend hoe snel na het begin van een spraakuiting, en op basis van welke geluidseigenschappen, we beslissen of we naar onze eigen taal luisteren of naar een vreemde taal. Wel is bekend dat we, ook zonder woorden te verstaan, alleen al aan de hand van de *zinsmelodie* kunnen vaststellen dat we naar een vreemde taal luisteren [12, en referenties aldaar].

Voorts zijn mensen in staat om heel nauwkeurig te bepalen waar geluiden vandaan komen, zowel naar richting als naar afstand. Bij het normale spraakverstaan speelt dit vermogen geen merkbare rol. Toch is gemakkelijk aan te tonen dat we een zwaar beroep doen op dit vermogen als we de stemmen van meerdere sprekers tegelijk uit elkaar moeten houden. Wie wel eens een monobandopname (dus zonder richtingsinformatie) van een discussie heeft moeten uitschrijven, weet uit ervaring hoe moeilijk dit wordt zodra twee of meer sprekers door elkaar heen praten.

Meer in het algemeen zijn wij mensen onbegrijpelijk knap in wat wel *selectief luisteren* genoemd wordt. We zijn in staat de spraak van een spreker te blijven volgen onder de meest barre omstandigheden, ondanks de aanwezigheid van achtergrondlawaai, stoorgeluiden en andere sprekers [15]. Er zijn zelfs situaties waarin we spraak beter kunnen verstaan in achtergrondlawaai dan in stilte.

Van deze vermogens begrijpen we nog maar weinig. Natuurlijk zouden we er graag veel meer van willen weten, niet alleen uit nieuwsgierigheid, maar ook omdat we dan wellicht met meer succes kunnen toewerken naar machines en computersystemen die onder realistische omstandigheden menselijke spraak kunnen verstaan.

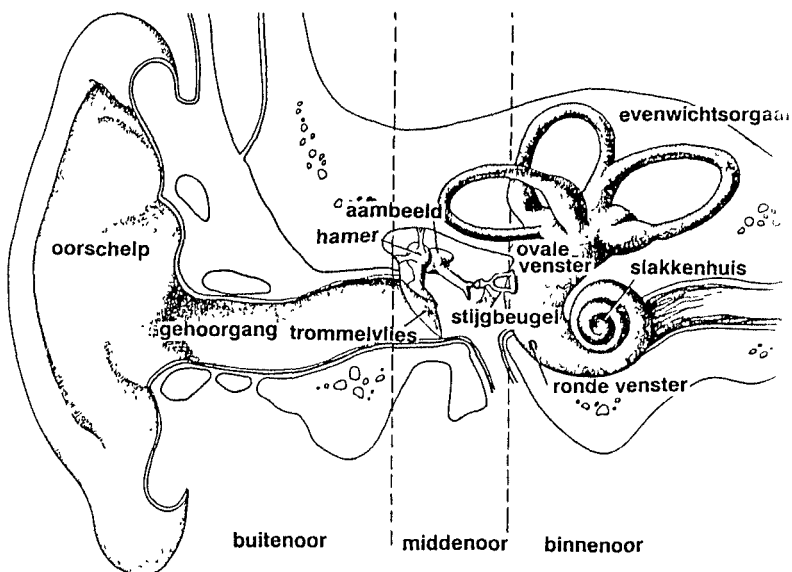
## 2. HET OOR

Het is gebruikelijk het oor onder te verdelen in drie stukken, het *buitenoor*, het *middenoor* en het *binnenoor*. Zoals duidelijk is uit de tekening in Figuur 1, omvat het buitenoor de *oorschelp*, en de *gchootgang*, die loopt tot aan het *trommelvlies*. De oorschelp helpt bij het richting horen.

Het middenoor is de ruimte die zich bevindt tussen het trommelvlies aan de buitenkant en het ovale venster meer hoofdinwaarts. De luchtdrukverschillen duwen het trommelvlies beurteelings naar binnen en zuigen het naar buiten. Via drie minuscule botjes, die naar hun uiterlijk *hamer*, *aambeeld* en *stijgbeugel* genoemd zijn, wordt de beweging van het trommelvlies doorgegeven aan het ovale venster. Omdat het ovale venster een veel kleiner oppervlak heeft dan het trommelvlies en omdat de botjes als een hefboom werken, worden de zwakke bewegingen van het trommelvlies mechanisch versterkt. De hamer kan in zijn bewegingen geremd worden door kleine spiertjes die reflexmatig worden aangespannen wanneer het oor getroffen wordt door heel luide, lage trillingen. Na bijvoorbeeld een rockconcert blijven we enige tijd hardhorend omdat deze zelfbeschermingsreflex zich niet meteen ontspant. Helaas komt de reflex te langzaam op gang om ook gehoorbeschadiging als gevolg van knallen en explosies te voorkomen.

Pas echt interessant wordt het in het binnenoor. Het binnenoor, dat om voor de hand liggende redenen (zie Figuur 1) ook wel het *slakkehuis* wordt genoemd, is aan de boven-

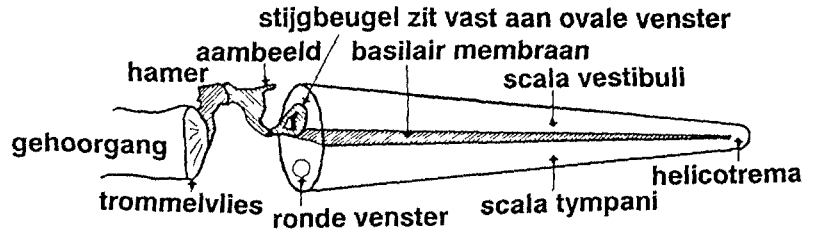
zijde getooid met drie halfcirkelvormige buizen. Hierin zetelt ons *evenwichtsorgaan*, dat het horen verder geen rol speelt. Als we het slakkehuis ontrollen (zie Figuur 2), zien we een buis van ongeveer 35 mm lengte, die naar de buitenzijde toe (aan de *basis*) relatief breed is, maar naar binnen toe taps toeloopt.



Figuur 1: Structuur van het *perifere gehoor*, met *buitenoor*, *middenoor* en *binnenoor*.

Het slakkehuis, dat gevuld is met vloeistof, wordt over zijn totale lengte horizontaal in tweeën gedeeld door het *basilair membraan*. Het membraan loopt echter niet helemaal recht, tot in de spits, waardoor de vloeistof in de bovenste kamer van het slakkehuis in verbinding staat met die in de onderste kamer. De vloeistof in het slakkehuis zit opgesloten tussen twee vliezen, het *ovale venster* aan de basis, en het *ronde venster* aan het uiteinde. De bewegingen van het ovale venster zetten zich voort als drukgolven in de vloeistof van het binnenoor. Na korte tijd gaat het basilair membraan meedeinen op de golfbewegingen van de vloeistof. Het blijkt nu dat de plaats van maximale deining meer naar de basis of naar het uiteinde van het basilair membraan komt te liggen naarmate de frequentie van de geluidsstimulus hoger, resp. lager is. Daarnaast leidt sterkere geluidsstimulering tot deining over een groter gebied van het basilair membraan.

stuk van het basilaire membraan. Op deze manier worden de eigenschappen frequentie en intensiteit van akoestische trillingen omgezet in eigenschappen van het bewegingspatroon van het basilaire membraan (zie verder Figuur 3)



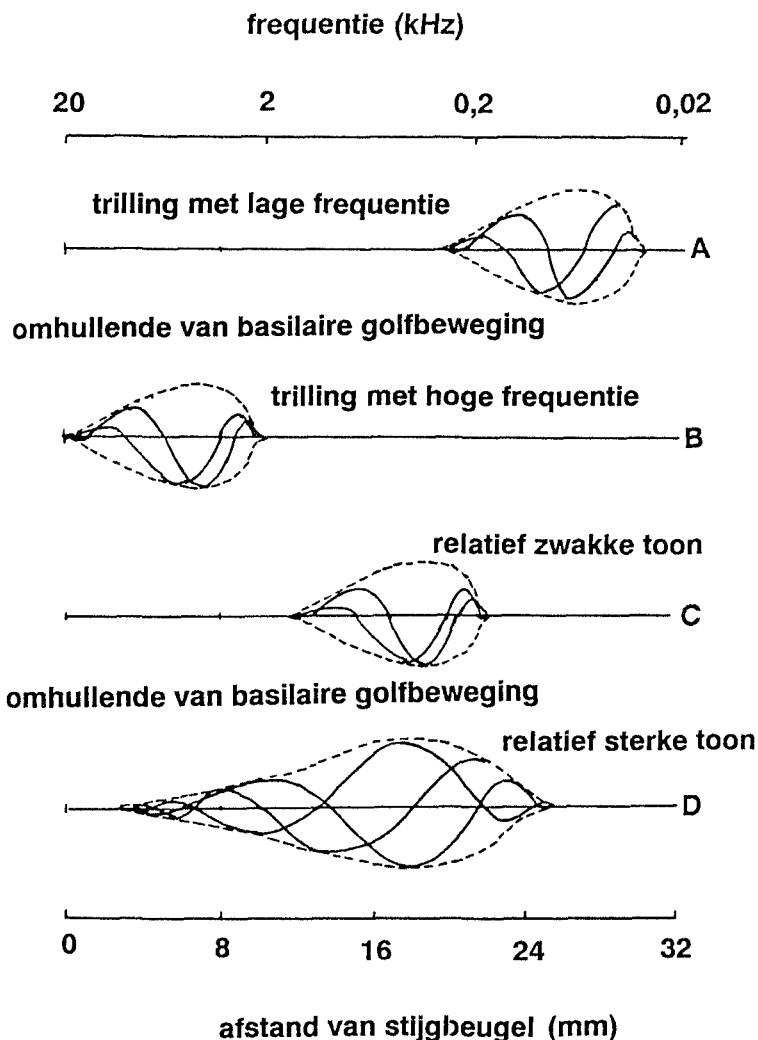
Figuur 2 Schematische voorstelling van het slakkehuis in (denkbeeldige) ontrolde toestand

Minstens drie eigenschappen van het basilaire bewegingsgedrag zijn belangrijk bij het spraakverstaan

Bij iedere verdubbeling van de akoestische frequentie komt de plaats van maximale uitwijking op het basilaire membraan ongeveer 3 mm dicht naar het ovale venster toe te liggen. De hoorbare eigenschap die we toekennen aan verschillen in frequentie noemen we toonhoogte. Algemeen geldt dat een hogere frequentie wordt waargenomen als een hogere toonhoogte. Voor het menselijk oor is een verdubbeling van de frequentie (een octaafsprong) altijd een even grote toonsprong, of het nu een verandering van 100 naar 200 Hz betreft of van 1000 naar 2000 Hz. Ons oor evalueert frequentiever verschillen dus als verhoudingen, ofwel logaritmisch.

Het stuk van het basilaire membraan dat meegolft is niet symmetrisch gespreid rond de plaats van de maximale uitwijking, maar ligt vooral aan de zijde waar zich het ovale venster bevindt, naar het gebied dus dat gevoelig is voor hoge tonen. Deze eigenschap zal ons straks in staat stellen asymmetrie in *toonmaskeringsverschijnselen* te begrijpen.

De basis (begin, bij het ovale venster) van het basilaire membraan is relatief dik en stijf. Door zijn mechanische eigenschappen kan dat deel van het membraan alleen meedineren op snelle golfbewegingen, d.w.z. hoge tonen. Door de dikte en stijfheid van dit stuk van het membraan zal de deining meteen ophouden als het geluid weg is. Het uiteinde van het membraan (bij de spits) is echter dun en flexibel, en is daardoor juist gevoelig voor langzame bewegingen, dus voor lage tonen. Bovendien zal dit deel van het membraan lang nadeineren na beëindiging van de laagfrequente geluidsstimulering. Aan de hand hiervan kunnen we straks begrijpen waarom *namaskering* door lage tonen langer duurt dan door hoge tonen.



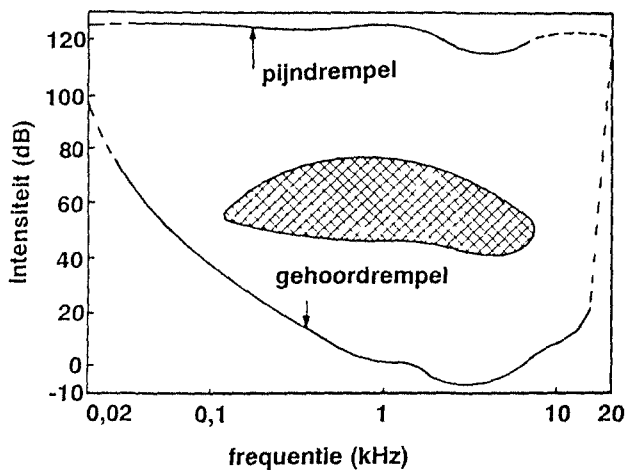
*Figuur 3:* Lopende golven op het basilair membraan als reactie op stimulering door een zuivere toon. A: stimulerende toon heeft een lage frequentie. B: stimulerende toon heeft een hoge frequentie. C: stimulerende toon heeft een frequentie in het middenbereik met relatief geringe sterkte. D: als C, maar nu met grote sterkte.

Over de hele lengte van het basilaire membraan bevinden zich zgn. *haarcellen*, een kleine 30 000 stuks. Elke haarcel geeft aan of het stukje van het basilaire membraan waarop deze zich bevindt, deint. Zo ja, dan geeft hij een zenuwimpuls af naar de hersenen (via de zenuwvezels die in de centrale gehoorzenuw gebundeld zijn), en blijft dat regelmatig doen totdat de deining ophoudt. Op deze manier kunnen onze hersenen precies bijhouden waar langs het basilaire membraan deining ontstaat, en dus ook welke frequenties in het geluid aanwezig zijn. Bovendien kunnen de hersenen uit de spreiding van de reagerende haarcellen uitrekenen hoe sterk een bepaalde frequentie is. Over de nauwkeurigheid waarmee we frequentie, geluidssterkte en -duur kunnen horen, zullen we het hierna hebben.

### 3. GELUIGHEID VAN HET GEHOOR

#### 3.1 ABSOLUTE HOORBAARHEID

Niet iedere luchtdrukverandering wordt door ons gehoor waargenomen als geluid. Als een drukverandering te gering is horen we niets, als de verandering te sterk is ervaren we pijn, maar geen geluid. Evenmin kunnen we heel langzame of heel snelle veranderingen als geluid waarnemen. Het gebied van hoorbare luchtdrukveranderingen ligt tussen frequenties van 20 tot 20 000 Hz, en intensiteiten van 0 tot 120 decibel (dB), zoals aangegeven in Figuur 4.



Figuur 4 Het gebied van hoorbare geluiden. Bij spraakgeluid ligt de intensiteit tussen 40 en 80 dB en liggen de relevante frequenties tussen 100 en 8000 Hz (zie het gearceerde gebied).

De notie *decibel* verdient toelichting. *Geluidsdruk* wordt gemeten in aantal dynes per  $\text{cm}^2$ . Voor een referentietoon van 1000 Hz is de geringste geluidsdruk die we kunnen onderscheiden van stilte  $0,0002 \text{ dyne/cm}^2$ , terwijl de krachtigste geluidsdruk een miljoen

keer sterker is 200 dyne/cm<sup>2</sup>. Bij afspraak stellen we de druk van het zachtste geluid we kunnen horen (de spreekwoordelijke vallende speld) op 0 dB. Een vertienvoudiging in de druk noemen we een relatieve toename van 2 Bel (ofwel 20 decibel). Het verschil tussen het zachtste en het luidste hoorbare geluid kunnen we dan overbruggen in 6 maal vertienvoudiging (dus een vermiljoenvoudiging), ofwel 6 maal 20 dB = 120 dB. De pijngrens ligt dan even boven de 120 dB. Onderstaande rijtjes illustreren een en ander.

geluiddruk in dyne/cm <sup>2</sup>	hoeveel maal sterker dan referentie	sterkteverschil t o v referentie in decibel	voorbeeld van geluid met deze sterkte
0,0002	1 x	0 dB	vallende speld
0,002	10 x	20 dB	fluisteren
0,02	100 x	40 dB	stad bij nacht
0,2	1000 x	60 dB	conversatie
2	10000 x	80 dB	schreeuwen
20	100000 x	100 dB	pneumat hamer
200	1000000 x	120 dB	straalmotor op 10

Voor ons gehoor is elke toename van de geluiddruk met bijvoorbeeld 10 dB een grote toename. Tevens blijkt een verschil in druk van 1 dB ruwweg het kleinste verschil te zijn dat we kunnen horen. Dit maakt de decibel tot een handige en zinvolle eenheid.

In Figuur 4 zien we dat ons oor voor frequenties tussen 1000 en 4000 Hz maximaal gevoelig is. Voor hogere, en vooral lagere, frequenties is ons oor een stuk minder gevoelig. Een toon met een frequentie van 1000 Hz kunnen we inderdaad bij 0 dB van stilte onderscheiden, maar om een toon van bijvoorbeeld 100 Hz te kunnen onderscheiden van stilte, moet hij al gauw een 50 dB sterker zijn.

Gelukkig liggen de frequenties die voor de spraakwaarneming belangrijk zijn precies in het gevoeligste deel van ons gehoorbereik. Bovendien spreken wij met een geluiddruk die keurig het midden houdt tussen de vallende speld en de pijngrens.

### 3.2 HOORBAARHEID VAN VERSCHILLEN

Er is in de loop van de jaren een uitgebreide literatuur opgebouwd over de gevoeligheid van het menselijk gehoor. Van allerhande eigenschappen van geluid is bekend hoe een verandering daarin minimaal moet worden aangebracht om te kunnen horen dat er iets veranderd is. Zo'n kleinste mogelijk, maar toch hoorbaar verschil heet een *verschilgrens* of *JWV* (*Just Waarneembaar Verskil*, ook wel *JND* naar het Engels *Just Noticeable Difference*). Zo zagen we al dat een verandering van de geluiddruk van 1 dB net genoeg is om te kunnen horen dat er een verschil bestaat tussen twee geluiden. Op dezelfde manier kunnen we ons afvragen hoe groot een verschil in frequentie tussen twee tonen minstens moet zijn willen we die tonen als verschillend ervaren, of hoe groot een verschil in duur, enz. We willen dan eerst weten welke eigenschappen van het geluid van belang zijn bij de spraakwaarneming, en vervolgens vaststellen hoe gevoelig ons oor is voor veranderingen van deze eigenschappen.

In 1 HART hebben we kunnen zien dat spraakgeluid nooit bestaat uit een enkelvoudige toon. Spraakklanken zijn altijd complexe, samengestelde trillingen, al dan niet met een periodieke structuur. Bovendien hebben we gezien dat de verdeling van de energie over het frequentiespectrum bij spraakklanken niet uniform is. In de literatuur is het meest onderzoek beschreven naar de waarneming van eenvoudiger gestructureerde geluiden, zoals enkelvoudige fluittonen en witte ruisstoten. Specifiek op spraakklanken aangelegd onderzoek is veel schaarser, daar valt nog veel te doen. Niettemin kunnen we, met de nodige slagen om de arm, wel wat voorlopige cijfers noemen.

Om enigszins betrouwbaar te kunnen horen dat er een toonhoogteverschil bestaat tussen twee losse, kunstmatige spraakklanken, moet het verschil in grondfrequentie minstens 0,3 tot 2,5% bedragen [23].

Een verschil in geluidsdruk moet ten minste 1 dB zijn om gehoord te worden.

Om te horen dat twee klanken verschillen in duur moet er een duurverschil zijn van ten minste 10% [5]. Er zijn aanwijzingen dat duurverschillen tussen stemloze klanken gemakkelijker te horen zijn dan tussen stemhebbende klanken [21]. Een nog nauwelijks onderzocht vraag is hoe nauwkeurig we de duur kunnen horen van stukjes stilte tussen spraakgeluiden in [8, 21].

Om te horen dat twee klinkers verschillen in *timbre* moet de middenfrequentie van de eerste of de tweede formant ten minste 3% verschillen [19, 34]. Wanneer twee of meer formanten tegelijk verschillen kan het verschil voor elke formant apart kleiner zijn, maar hoeveel precies is niet bekend [31]. Men kan vermoeden dat de resultaten bij gesproken klinkers anders liggen dan bij gefluisterde klinkers. Ook hierover is (mij) niets bekend.

Tot nu toe hebben we het gehad over eigenschappen van niet-veranderende geluiden. In spraak veranderen spraakklanken voortdurend van toonhoogte, timbre en sterkte. Met name het onderzoek naar de gevoeligheid waarmee we de snelheid en richting van veranderingen binnen spraakklanken waarnemen, staat nog helemaal aan het begin. Toch enkele eerste indrukken.

Om te horen dat twee accentverlenende grondtoonbewegingen van elkaar verschillen in spronggrootte, is een verschil van 1,5 semitoon (ca. 9%) voldoende [23, 37].

Om te horen dat twee klanken verschillen in de mate van abruptheid waarmee ze inzetten of eindigen, moet de amplitude aan het begin of eind van de klanken ten minste 25% sneller of langzamer oplopen of dalen [14, 25].

Aan de hand van de waargenomen duren van herkende klanken en woorden kan de luisteraar het spreektempo schatten. Er is nog maar heel weinig bekend over hoe nauwkeurig verschillen in spreektempo worden waargenomen, of op grond waarvan een luisteraar hoort dat een spreker zijn tempo versnelt of vertraagt [10] (zie verder DLN 05).

Er is een begin gemaakt met onderzoek naar de gevoeligheid waarmee we verandering van grondfrequentie en verandering van formantfrequenties waarnemen. Hieruit blijkt o.a. dat kortdurende veranderingen moeilijk hoorbaar zijn, en dat over het algemeen frequentieveranderingen vrij groot moeten zijn voordat we überhaupt horen dat het geluid niet langer stationair blijft. Een slotte blijkt dat, als we al kunnen horen dat





frequenties veranderen, het heel lastig is om te horen of de ene verandering sneller dan de andere, of zelfs of de verandering een stijging of een daling van de frequentie inhoudt [35]

### 3.3 MASKERINGSVERSCIJNSLIJN

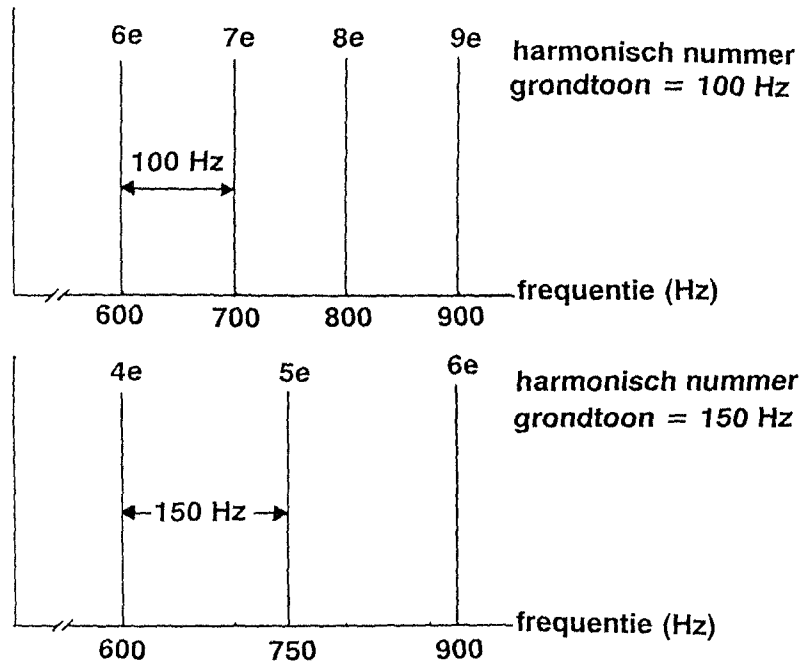
Omdat zelfs een zuivere toon al leidt tot activering van haarcellen over een groot gebied van het basilaar membraan, is het oor niet goed in staat andere in frequentie naburige tonen waar te nemen. We zeggen dan dat de ene toon de andere maskeert, d.w.z. onhoorbaar maakt. Ieder geval minder hoorbaar maakt. Over het algemeen zien we dat een lage toon een hogere toon sterker maskeert dan omgekeerd. Het gebied waarbinnen tonen elkaar de waarneembaarheid beïnvloeden, noemen we een *kritieke band*. Heel grof gesteld het gebied van gelijktijdige tonen list van elkaar zolang zij minder dan een kleine terts van elkaar verschillen: hun frequentieverhouding is dan kleiner dan 1,125.

We hebben eerder kunnen vaststellen dat ons gehoor, binnen het spraakgebied, moeite heeft lage tonen te horen dan hoge (zie Figuur 4). Het asymmetrische maskeringsgedrag daarentegen bevoordeelt lage tonen boven hoge, zodat de balans van de frequentieszins hersteld wordt.

De menselijke stembandtrilling bevat een grondtoon en in principe alle bovengrondtonen. Boventonen hebben frequenties op hele veelvouden van de grondtoon. De eerste bovengrondtoon ligt precies een octaaf boven de grondtoon, en valt dus buiten diens kritieke band. Hoe bovengrondtonen komen verhoudingsgewijs steeds dichterbij elkaar te liggen, en hoe (ruwweg) de 12e bovengrondtoon zijn we niet meer in staat individuele bovengrondtonen in te horen. In een complexe toon 'uit te luisteren', d.w.z. apart waar te nemen. Naastliggende bovengrondtonen vallen dan volledig binnen elkaars kritieke banden. Geluiden maskeren elkaar sterker naarmate hun stimuleringsgebieden op het basilaar membraan elkaar meer overlappen.

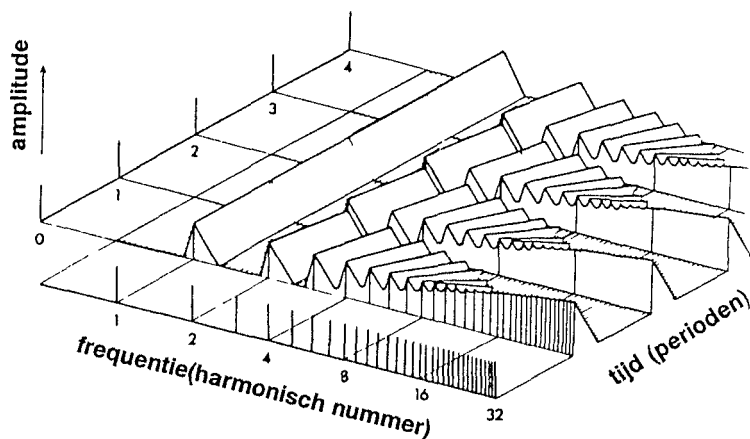
De toonhoogte van een tooncomplex, zoals van een menselijke klinker, leiden we af uit de ligging van de lagere, apart hoorbare harmonischen. Geen enkele individuele harmonische, ook de grondtoon niet, bepaalt op zich de waargenomen toonhoogte. Het menselijk gehoor rekent de toonhoogte uit aan de hand van de afstanden tussen plussen van maximale stimulering op het basilaar membraan als gevolg van apart waargenomen bovengrondtonen. De toonhoogte in normale spraakgeluiden is dan gelijk aan de grootste gedeelte van de waargenomen frequenties (zie Figuur 5).

Niet alleen gelijktijdig klinkende geluiden kunnen elkaar maskeren. Als we 's nachts in de auto een tegenligger ontmoeten die vergeet zijn koplampen te dimmen, dan zijn wij enige tijd verblind. In die periode zijn onze ogen niet in staat zwakke lichtschijnsels waar te nemen. Op dezelfde manier kunnen onze oren korte tijd verdoofd raken na het horen van een krachtig geluid. We noemen dit *namaskering* of *voorwaartse maskering*. In normale omstandigheden duurt deze tijdelijke ongevoeligheid hooguit 200 ms. Daarna heeft het gehoor zich voldoende hersteld om ook weer de details van zwakke geluiden waar te nemen. In dit opzicht komt het heel goed uit dat zgn. plofklanken (bijvoorbeeld [p] of [d]) een korte stilte bevatten op de overgang tussen de voorgaande (luide) klinker (veel zwakkere) explosie van de medeklinker zelf. Laboratoriumproeven hebben uitgewezen dat het verschil tussen bijvoorbeeld [p], [t] en [k] een stuk moeilijker te horen is wanneer we die natuurlijke stilte uit het spraakgeluid elimineren [39, 43].



*Figuur 5: Het gehoor bepaalt de toonhoogte in een tooncomplex door de grondtoon uit te rekenen die het best past bij de hoorbare harmonischen. In normale gevallen is dat de grootste gemene deler van de boventoonfrequenties. A: bij een tooncomplex met harmonischen van 600, 700, 800 en 900 Hz (resp. de 6<sup>e</sup>, 7<sup>e</sup>, 8<sup>e</sup> en 9<sup>e</sup> harmonische van een grondtoon van 100 Hz). B: bij een tooncomplex met harmonischen bij 600, 750 en 900 Hz (resp. de 4<sup>e</sup>, 5<sup>e</sup> en 6<sup>e</sup> harmonische van een grondtoon van 150 Hz).*

Het gehoor herstelt zich sneller van een hoogfrequent geluid dan van een lage trilling. Met name in het gebied onder de 400 Hz klinken tonen lang na in ons gehoor. Dit is mogelijk ook een (gedeeltelijke) verklaring voor het feit dat we ons niet bewust zijn van onderbrekingen van de zinsmelodie bij stemloze medeklinkers. *Figuur 6 illustreert treffend de besproken beperkingen van het oplossend vermogen van ons oor in tijd en frequentie.*



*Figuur 6* Geidealiseerde voorstelling van de reactie van het basilaire membraan op stimulering door een periodieke puls. De tijd loopt lineair van voor naar achteren, de frequentie logaritmisch van links naar rechts en de amplitude staat verticaal uit langs de logaritmische as. Uiterst links de golfvorm van de pulserieks, op de voorgrond het bijbehorende spectrum. Het midden van de Figuur laat de (denkbeeldige) omhullende van de lopende golf langs het basilaire membraan zien als functie van de tijd. [N. Duifhuis, H. (1972) *Perceptual analysis of sound*. Dissertatie TH Eindhoven]

In recente pogingen de prestaties van systemen voor *automatische spraakherkenning* verbeteren, streeft men ernaar de eigenaardigheden van het oor na te bootsen. Door aan het begin van de automatische spraakherkenner een oormodel te plaatsen (een zgn. *front processor*) kan het herkenningmechanisme zich meteen concentreren op de essentiële eigenschappen van het binnenkomend geluid, en zo sneller en efficiënter beslissen welk woord in het geheugen daar het meest op lijkt. Zo'n oormodel als voorzetstuk blijkt voor een verbetering op te leveren bij automatische spraakherkenning in achtergrondlawaai (vgl. POLS).

## 4 WAARAAN HERKENNEN WE DE DIVERSE SPRAAKKLANKEN?

## 4.1 IS SPRAAKVERSTAAN HETZELFDE ALS KLANKEN VERSTAAN?

Het is verleidelijk ons voor te stellen dat we bij het spraakverstaan de geluidsstroom die ons oor binnenkomt opdelen in stukjes ter grootte van aparte klanken. Klanken zijn, in deze visie, stukjes geluid die we in gedrukte tekst terugvinden als letters. We beslissen dan bij iedere klank onmiddellijk welke het is. Daarbij kijken we steeds in het woordenboek in ons achterhoofd (LAMMENS), om te zien of de reeks geïdentificeerde klanken al overeenkomt met een woord. Als dat zo is, dan hebben we een woord herkend. Reeksen herkende woorden brengen we vervolgens met elkaar in verband, en zo begripen we zinnen.

Deze primitieve voorstelling is in de praktijk niet houdbaar. Het is vrijwel uitgesloten dat we bij het normale spraakverstaan aparte klanken kunnen identificeren, om daarna pas woorden te herkennen.

Bij het spreken bewegen onze spraakorganen (mond, lippen, tong, enz.) zich volmaakt vloeiend van de ene klank naar de volgende. Wat we klanken noemen, zijn in feite denkbeelden in de geest van de spreker en de luisteraar, in de spraak zelf hebben we alleen maar bewegingen. Als gevolg hiervan verandert ook het spraakgeluid voortdurend. Het is daarom onmogelijk voor een luisteraar om te zeggen: 'hier is een klank afgelopen, en begint de volgende'. Zolang de luisteraar niet weet van waar tot waar klanken lopen, kan hij natuurlijk ook niet zeggen welke klanken hij hoort.

Als we in laboratoriumproeven spraakklanken - zo goed en zo kwaad als dat gaat - losnijden uit de woorden waarin ze gesproken zijn, dan blijkt zo'n geïsoleerde klank meestal niet herkenbaar te zijn. De spraakklanken zijn stuk voor stuk te slordig uitgesproken om ze te kunnen identificeren. Pas als meerdere klanken met elkaar een gekende eenheid vormen, hebben we voldoende duidelijkheid over de identiteit van elke aparte klank.

Hoe mensen woorden herkennen zullen we hier verder niet bespreken. Daaraan zijn aparte hoofdstukken (NOOTLBOOM, QUENÉ) gewijd in dit boek. Evenmin zullen we echt kunnen ingaan op de manier waarop mensen uit een reeks woorden een zinsbetekenis destilleren. Wel is gebleken dat zinsmelodie en fraseering een belangrijke bijdrage leveren aan de verwerking van zinnen, zie verder de hoofdstukken van DEN OS (over *ritmiek*), RIETVELD (over *klemtoon en accent*) en COLLIER (over *zinsmelodie*). Aan het eind van dit hoofdstuk zal ik enige aandacht besteden aan enkele problemen rond de spraakwaarneming die op dit ogenblik het wetenschappelijk veld bezighouden. Voor het overige zullen we ons in dit hoofdstuk een beetje van den domme houden, en doen alsof we klanken succesvol kunnen herkennen in eenheden die niet groter zijn dan een lettergreep. Hierover is na zo'n 20 jaar intensief onderzoek, met name in de Verenigde Staten, veel bekend.

## 4.2 KLINKER EN GLOVER MEDEKLINKER

Zoals we in eerdere hoofdstukken hebben kunnen zien, is het gebruikelijk spraakklanken in te delen in categorieën. Het bekendste onderscheid is dat tussen klinkers en medeklinkers. Binnen de groep medeklinkers kunnen we klanken verder uitsplitsen naar bijvoorbeeld de

manier waarop de uitstromende lucht in de mond en/of keelholte gehinderd wordt, en de belangrijkste hindernissen zitten. We zouden nu graag willen weten aan de hand welke eigenschappen van het spraakgeluid de luisteraar beslist met wat voor spraakklank hij te maken heeft. Hoe weet de luisteraar of hij een klinker of medeklinker heeft gehoord?

Bij klinkers fungeert het hele keel-mondkanaal als *verkleurend filter* (zie 'T II). Daardoor zitten er in de spectrale omhullende van een klinker goed-gemarkeerde topjes (*formanten*) op min of meer regelmatige afstanden. Wat we precies moeten verstaan onder goed-gemarkeerd is nog onduidelijk. Laten we zeggen dat de *bandbreedte* van de vormant gering moet zijn t.o.v. de middenfrequentie. Als we de toppen erg breed maken, tegenwoordig in het laboratorium een koud kunstje is, dan kunnen we de verschillende klinkers niet meer van elkaar onderscheiden. Als we daarentegen spraak maken door alleen maar een zuivere toon te genereren op de plaats van de middenfrequenties van de eerste drie formanten (drie formanten dus met een oneindig kleine bandbreedte), dan is het resultaat - met enige moeite - nog steeds te verstaan (zgn. *sinusgolf analogons*) [36].

Bij een gemiddelde mannenstem zitten in het gebied tot 5000 Hz bij neutrale tonus ongeveer vijf van zulke toppen, en wel bij 500, 1500, 2500, 3500 en 4500 Hz. De verschillende klinkers in de taal maken we dan door de ligging van deze vijf toppen (en met name van de laagste twee of drie) wat t.o.v. elkaar te verschuiven.

Essentieel voor het waarnemen van een klinker is dat ons oor duidelijke topjes signaleert. Overigens zijn er ook medeklinkers met een goed-gemarkeerde formantstructuur, de sonorante medeklinkers zoals [w], [j], [l], [r] en de nasalen. In zulke gevallen berust het onderscheid tussen klinker en medeklinker vooral op een verschil in geluidsterkte. Klinkers hebben een grotere intensiteit dan deze medeklinkers. Hieruit blijkt al dat het vrijwel ondoenlijk is van één enkele klank te zeggen of het een klinker dan wel een medeklinker is. We moeten minstens een klinker en een medeklinker naast elkaar hebben, doorgaans een lettergreep dus, voordat we kunnen zeggen wie wat is.

Ook zijn klinkers bij uitstek *stemhebbend*, terwijl echte medeklinkers het gemakkelijk *stemloos* gemaakt worden. Aan stemhebbende klanken kan ons oor een *toonhoogte* toegekennen. Er zijn *boventonen*, d.w.z. uitwijkingen op regelmatige (logaritmisch aflopende) afstanden langs het basilaar membraan (zie boven). Stemloze klanken hebben geen eenvoudig definieerbare toonhoogte. Toch is dit onderscheid van ondergeschikt belang. Als we fluisteren trillen de stembanden helemaal niet, terwijl uitstekend te horen blijft wat de klinkers en wat de medeklinkers zijn.

We zien hier een belangrijk gegeven in de waarneming van spraakklanken. Klinkers onderscheiden zich altijd van elkaar in meerdere opzichten tegelijk. We zullen aanstonds nog diverse malen tegenkomen. Het is dan interessant te leren op welke verschillen het oor het meest acht slaat, en na te gaan of verschillen die normaal ondergeschikte rol spelen, onder speciale omstandigheden doorslaggevend kunnen worden.

### 4.3 VERSCHILLEN TUSSEN KLINKERS

De verschillende klinkers herkennen we in de eerste plaats aan de ligging van de formanten in het spectrum. Het gaat hierbij vooral om de plaats van de middenfrequenties van de eerste drie formanten. De bandbreedtes luisteren niet zo nauw, vooropgesteld dat deze klein genoeg zijn om überhaupt de gedachte aan een klinker op te roepen. Klinkers kunnen in

benadering goed geïdentificeerd worden op basis van hun laagste drie formanten. De ligging van de  $F_1$  correspondeert hierbij vooral met de graad van mondopening waarmee de klinker tot stand is gekomen. Gesloten klinkers zoals [i] of [u] hebben een relatief lage  $F_1$ , voor een mannenstem in de buurt van 250 Hz. Open klinkers als [a] hebben een hoge  $F_1$  van ongeveer 800 Hz. Aan de ligging van de  $F_2$  kan het oor vaststellen of het te maken heeft met een voorklinker of een achterklinker. Hoe verder naar voren in de mond de klinker gearticuleerd wordt, des te hoger komt de  $F_2$  te liggen. Voor een [i] kan de  $F_2$  bij 2300 Hz liggen, voor een [u] bij 600 Hz. Ook is wel gesuggereerd dat het voor/achterkarakter van een klinker het best beluisterd kan worden aan de afstand tussen de  $F_2$  en de  $F_1$ . Hoe groter deze afstand hoe verder naar voren een klinker gearticuleerd is. Als klinkers gemaakt worden met getuile lippen, bijvoorbeeld [u, y], dan wordt het spraakkanaal een stukje langer, waardoor met name de  $F_2$  en de  $F_3$  een lagere frequentie krijgen.

Hoe de klinkers op het gehoor het best van elkaar te onderscheiden zijn, wordt duidelijk gemaakt wanneer we de formantfrequenties niet uitdrukken in Hertz, maar kijken naar de plaats van de formanttoppen op het basilaar membraan. De afstanden tussen de formanten zetten we dan af langs een zgn. *Bark-schaal* [11, 29]. Een *Bark* is een afstand tussen twee tonen ter grootte van een kritieke band. Tonen die minder dan 1 Bark van elkaar af liggen, kan ons oor niet onafhankelijk van elkaar waarnemen (zie boven). Vergelijk de rangschikking van klinkers in de Bark-plot (Figuur 7b) met die in de Hertz-plot (Figuur 7a).

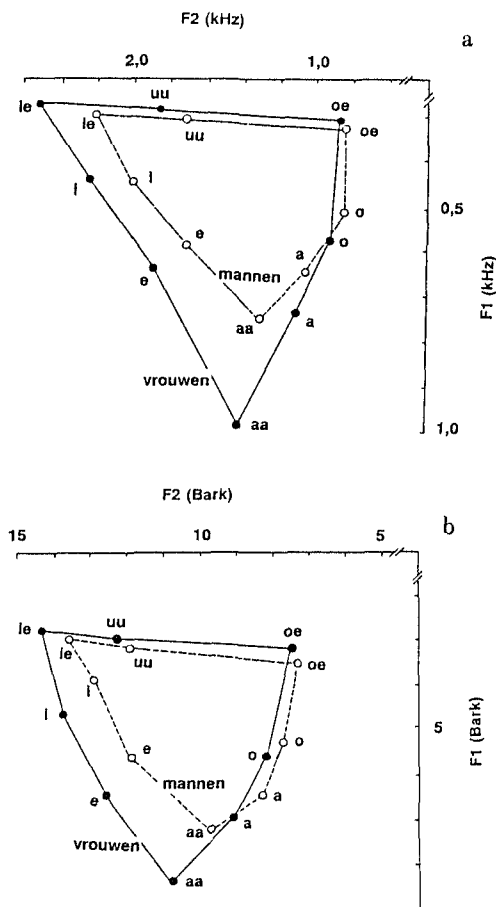
Een probleem apart vormen de *nasale klinkers* zoals we die tegenkomen in het Franse zinnetje *un bon vin blanc*. Bij nasale klinkers is er een opening vanuit de keel naar de neusholte. Daardoor ontstaat een extra formant met een heel lage frequentie, de zgn. *nasaalformant*. Tegelijkertijd echter worden hogere frequenties uitgedoofd, waardoor in het spectrum een deuk ontstaat, een zgn. *nul-* of *antiformant*. Als het gebied waarover tonen ontbreken maar breed genoeg is, is ons oor wel gevoelig voor zo'n deuk, en nemen we een nasale klinker waar.

Om te kunnen vaststellen of een klinker een tweeklank is, moet ons gehoor letten op een verandering in formantligging. Hierbij is ons oor niet al te selectief: als er maar een duidelijke en betrekkelijk langzame verandering optreedt in formantfrequenties in een bepaalde richting dan horen we een tweeklank. De richting van de verandering is hierbij belangrijker dan daadwerkelijk bereikte eindfrequentie(s).

Klinkers kunnen voorts systematisch verschillen in duur. In het Nederlands onderscheiden we korte en lange klinkers. In andere talen, bijvoorbeeld het Estisch, bestaan zelfs korte, lange en superlange klinkers [28].

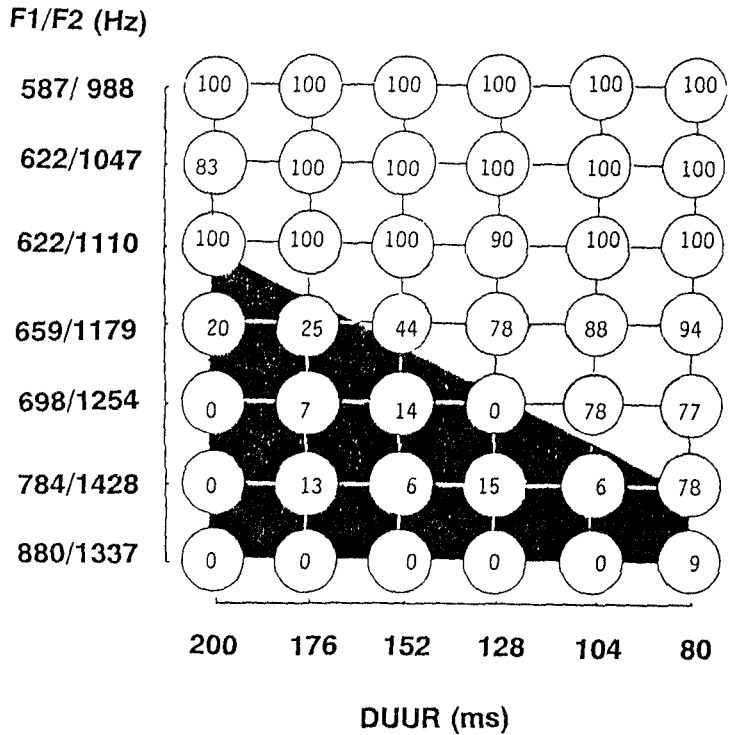
In het Nedeilandshe hebben we bijvoorbeeld een korte [α] en een lange [a]. De [a] duurt ongeveer twee keer zo lang als de [α]. Daarnaast echter articuleren we de [a] wat meer naar voren in de mond, die ook nog eens wat verder geopend wordt dan bij de [α], waardoor de  $F_1$  en de  $F_2$  van de [a] wat hoger liggen dan die van de [α]. Waaraan horen we nu of we met een [a] of met een [α] te maken hebben?

Uit Figuur 8 blijkt dat de duur- en de formantverschillen in ongeveer gelijke mate bijdragen aan het onderscheid. Bovendien blijkt er onderlinge compensatie mogelijk als een klinker op grond van zijn formantligging een [a] zou moeten zijn, kunnen we er in het laboratorium toch een [α] van maken door zijn duur extra kort te maken, en vice versa [26]. Deze onderlinge compensatiemogelijkheden hebben in de literatuur veel aandacht gekregen onder de naam *trading* (zie verder sectie 6.3).



**Figuur 7** Ligging van de Nederlandse niet-verglijdende klinkers in een F1/F2-veld (in formantfrequenties uit Pols, L.C.W. (1977). *Spectral analysis and identification of Dutch vowels in monosyllabic words*. Dissertatie VU Amsterdam). 7A: de klinkers uitgezet in lineaire frequentieassen in Hertz. 7B: de klinkers uitgezet na omrekening van formantfrequenties in perceptief relevante schalen m.b.v. de Bark-transformatie, wat een afstand van 1 Bark overeenkomt met 1 kritieke band. Merk op dat de klinkers na Bark-transformatie evenwichtiger gespreid liggen in het klinkerveld. Bovendien vertonen klinkersystemen van mannen (doorgetrokken lijnen) en van vrouwen (onderbroken lijnen) nu meer overeenkomst. Een handige formule voor de Bark-transformatie is  $f_{\text{bark}} = 7 * \ln(f_{\text{Hz}}/650 + \sqrt{(f_{\text{Hz}}/650)^2 + 1})$

$$t_{\text{bark}} = 7 * \ln(t_{\text{Hz}}/650 + \sqrt{(t_{\text{Hz}}/650)^2 + 1})$$



*Figuur 6* Percentage [a]-oordelen (complement is [α] oordelen) als functie van klinkerkwaliteit (F<sub>1</sub> en F<sub>2</sub> in Hz verticaal aflopend) en klinkerduur (in ms van links naar rechts aflopend). De scheidslijn tussen het [a]-gebied (zwart) en het [α]-gebied (wit) is de zog. *fonceme grens*.

De identificatie van klinkers is tot nog toe overdreven vereenvoudigd voorgesteld. Ik stel twee complicaties aan de orde.

De ligging van de formanten is sterk afhankelijk van de spreker. De formanten van een spreker met een groot hoofd, dus met grote resonantieholten, liggen bij lagere frequenties dan die van een spreker met een klein hoofd, bijvoorbeeld vrouwen en kinderen. In Figuur 7B komen de formanten dan allemaal 1 tot 2 Bark hoger te liggen. Wil ons gehoor kunnen beslissen welke klinker er gesproken wordt, dan moet het oor eerst weten wat voor een spreker er aan het woord is, een met een groot hoofd of een met een klein koppie. Hoewel er al veel onderzoek naar gedaan is [17], weten we nog steeds niet goed hoe ons gehoor in staat is de beeldvertekening als gevolg van sprekeverschillen te corrigeren [42]. Niettemin blijken luisteraars voortreffelijk in staat



de sprekergebonden eigenaardigheden te scheiden van de voor de spraakherkenning essentiële informatie in de geluidsstroom

Wanneer de klinkers niet langzaam en zorgvuldig worden uitgesproken, wordt de afstand tussen de klinkers in het klinkerveld kleiner. Toch blijft ons gehoor deze klinker horen, of die nu snel en slordig of langzaam en nauwkeurig gearticuleerd wordt. In ons brein voeren we kennelijk een rekensom uit, waarmee we deze verschillen kunnen wegwerken. Alweer: hoe we dit precies doen, is nog steeds onduidelijk (zie sectie 5.3)

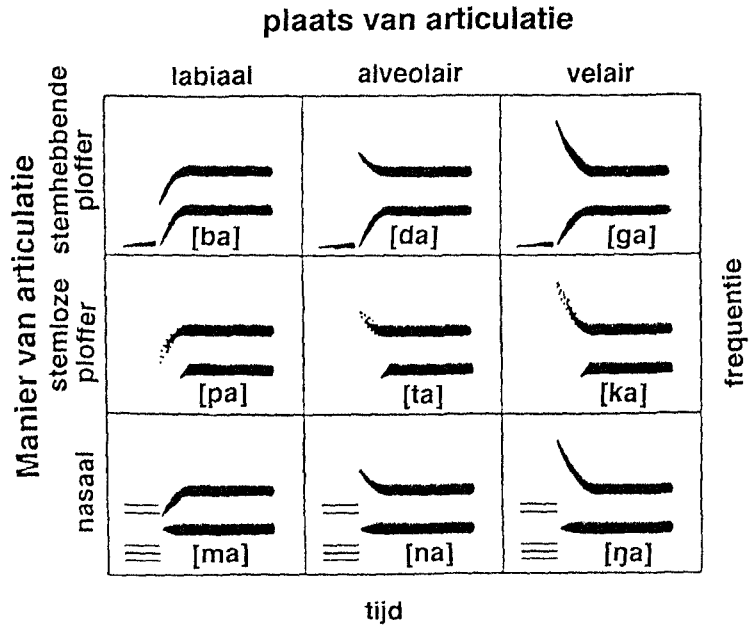
#### 4.4 VERSCHILLEN TUSSEN MEDEKLINKERS

In de categorie medeklinkers onderscheiden we allereerst de *obstruenten* (de medeklinkers) van de *sonoranten* (de klinkerachtige medeklinkers). *Echte medeklinkers* worden gemaakt met een vernauwing ergens in de mond of keel die of de uitstromende lucht geheel onderbreekt, waardoor een kort moment van stilte ontstaat, of leidt tot hoorbare luchtwevelingen, dus ruis. In beide gevallen ontbreekt de formantstructuur kenmerkend is voor klinkers. Obstruenten herkennen we dus aan het kortstondig wegvalen van formanten. Sonoranten bezitten wel zo'n formantstructuur, maar hun intensiteit is zwakker dan die van klinkers.

##### Obstruenten

Bij zijn *ploffers* is er (vrijwel) geen geluid in de tijd dat de formantstructuur afwezig is. De luchtdoorgang in de mond of keel is dan volledig geblokkeerd. Op het moment dat de blokkade wordt opgeheven ontstaat een korte explosie. Deze knal duurt ongeveer 10 ms. Na de stilte heel goed waar te nemen. Aan de kleuring van de knal is uitstekend te herkennen waar in de mond of keel de blokkade zich bevond, bij de lippen, op de tandkassen of op de zachte gehemelte. De knal bestaat in beginsel uit witte ruis, waarin alle tooncomponenten met gelijke sterkte vertegenwoordigd zijn. De holte die zich voor de afsluiting bevindt fungeert weer als verkleurend filter. Is deze voorste holte groot, zoals bij een veelgrote ploffer, dan is de ruis dof gekleurd, d.w.z. hij heeft overwegend lage frequenties. Is de voorste holte klein, zoals bij een tandkasklank [t], dan is de ruis veel scherper gekleurd. Lipklanken zit er in feite helemaal geen holte meer voor de afsluiting, waardoor de ruis uit deze klanken min of meer ongekleurd (wit) blijft. De verschillende kleuringen van de ruis leiden tot stimulering van verschillende delen van het basilaire membraan. In tegenstelling tot klinkers (formanten) is deze stimulering breedbandig.

Zo'n afsluiting van het mond-keelkanaal wordt tamelijk geleidelijk gemaakt. Onze tong of lippen doen er 30 tot 100 ms over om zich vanaf hun klinkerstand naar de medeklinkerpositie te bewegen en nog eens een keer zo iets om daarna weer terug naar de klinkerstand aan te nemen. Tijdens deze bewegingen verandert de vorm van het spraakkanaal relatief snel, waarbij de formanten stijgen of dalen in frequentie. Vooral de beweging van de  $F_2$  kunnen we goed horen waar in de mond of keel de afsluiting plaatsvindt (zie Figuur 9).

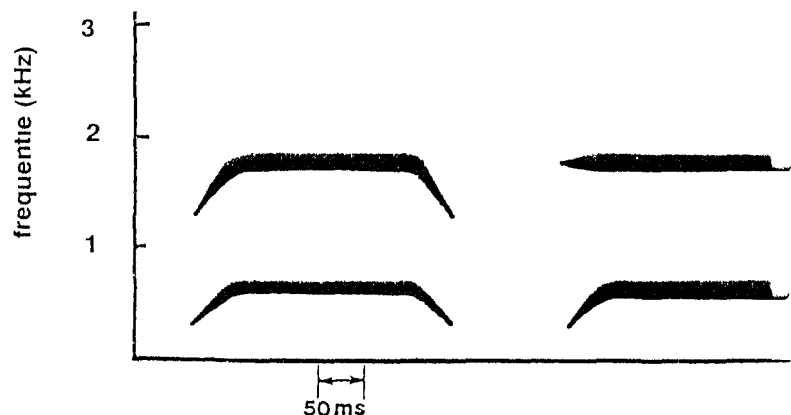


Figuur 9: Gestileerd verloop van  $F_1$  en  $F_2$  in enkele eenvoudige syllaben. Wanneer we volgens deze patronen een spraaksynthesator aansturen, ontstaan duidelijk herkenbare medeklinkers gevolgd door een klinker [a]. Merk op dat de plaats van articulatie vooral af te leiden uit het verloop van de  $F_2$ . Deze ontspringt voor een labiale klank bij 700 Hz en veegt dan in 50 ms naar de doelwaarde voor de klinker; bij een alveolaire medeklinker ontspringt de  $F_2$  bij 1700 Hz, en voor een velaire medeklinker bij 2700 Hz.

De kleur van de knal en de formantbewegingen voor en na de stilte zijn niet constant voor een bepaalde articulatieplaats. Zij worden mede beïnvloed door de eigenschappen van de ingrenzende klinker(s). Er is nog steeds veel onderzoek gaande om uit te maken hoe precies de articulatieplaats van een ploffer herkend wordt in verschillende klinkeromgevingen (zie verder [20] en verwijzingen aldaar, met name naar het werk van Stevens Blumstein).

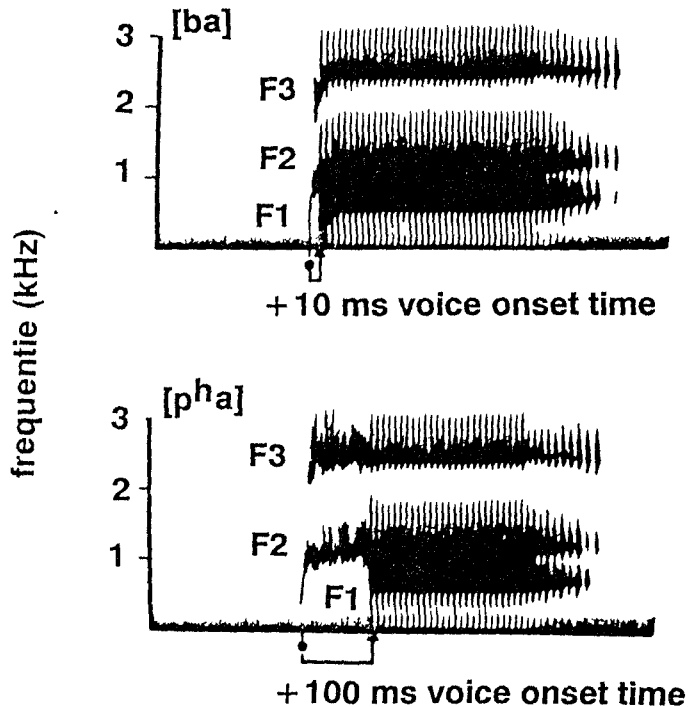
We hebben dus nu minstens drie aanwijzingen in het geluid die ons kunnen vertellen wat de articulatieplaats van een plofklank is: de kleur van de knal, de  $F_2$ -verandering (formantvergliding of -transitie) voor de stilte, en de  $F_2$ -verandering na de knal. In normale

spraak werken deze aanwijzingen samen, maar in kunstmatige spraak kunnen we ze in elkaar in conflict brengen. In zulke situaties blijkt de kleur van de knal doorslaggevend zijn. Als we de knal weglaten (wat in ieder geval gebeurt bij stemhebbende ploffers) en  $F_2$ -bewegingen voor en na de stilte met elkaar in conflict brengen, dan horen we niet meer maar twee medeklinkers achter elkaar: de eerste heeft dan de articulatieplaats die wordt gesuggereerd door de  $F_2$ -beweging naar de stilte toe, de tweede die van de  $F_2$  beweging na de stilte (zie Figuur 10).



Figuur 10 Gestileerd verloop van  $F_1$  en  $F_2$ , geschikt als recept voor herkenbare synthese van de syllabereeks [bebde]. In een medeklinkeropeenvolging als [-bd-] is de  $F_1$ -aanbuiging van de eerste syllabe bepalend voor de waargenomen articulatieplaats van de [-b] en de  $F_2$ -beginbuiging van de tweede syllabe bepalend voor de [d].

Het verschil tussen stemhebbende en stemloze ploffers horen we, in weerwil van de raamgeving, niet of nauwelijks aan de aan- of afwezigheid van stembandtrilling, of in andere termen - aan een *toonhoogtegewaarwording*. Het onderscheid blijft goed beluisterbaar in fluisterspraak, waar de stembanden überhaupt niet trillen. We hebben boven al gezien dat de bewegingen van het uiteinde van het basilaar membraan (bij lage tonen dus) niet meteen ophouden als het geluid verdwijnt. De aan- of afwezigheid van alleen maar wat  $F_1$ -tonen (stembandtrilling) in de stilte van een ploffer is dus op zich geen goede aanwijzing [32]. Wel is belangrijk dat de stilte, of afwezigheid van formantstructuur, bij stemloze ploffers veel langer duurt dan bij stemhebbende, dat de knal bij stemloze ploffers langer duurt en luider is dan bij stemhebbende, en dat de formantveranderingen voor en na de knal bij stemhebbende ploffers wat langzamer gaan dan bij stemloze [40]. In sommige talen (bijv. Engels) treedt onmiddellijk na de explosie een stemloos (d.w.z. *gefluisterd* of *aspirated*) klinkerdeel op. Hierbij verdwijnt de periodiciteit uit het spraakgeluid en ontbreekt de  $F_1$  (zgn.  $F_1$ -cutback, zie Figuur 11). Ook hiervan is aangetoond dat de waarneming van het stemhebbend/stemloos contrast erdoor beïnvloed wordt.



*Figuur 11* Onderlinge afhankelijkheid van *Voice Onset Time* (VOT), aspiratieduur, en  $F_1$ -utback in het Amerikaans-Engelse klankcontrast [ba] [pʰa]

In dit verband is in de literatuur vrij veel aandacht besteed aan het verschijnsel *steminzet tijd* *voice onset time*, of *VOT*. Als we het moment van de knal (de overgang van stilte naar geluid) bij ploffers als referentiepunt nemen, dan geldt het volgende. Bij stemhebbende ploffers (aan het begin van een spraakuiting) begint de stembandtrilling al enige tijd voor de knal (30 tot 100 ms) de steminzet tijd krijgt dan een negatieve waarde van -30 tot -100 ms. Stemloze ploffers hebben een steminzet tijd van 0 of hoger. Bij 0 valt de steminzet samen met de knal, bij een positieve waarde beginnen de stembanden pas te trillen nadat de explosie is afgelopen. Het tijdsinterval tussen de knal en de steminzet wordt dan opgevuld met *fluisteruis* (zgn. *aspiratie*), die verder dezelfde kleur heeft als het daaropvolgende stemhebbende klinkerdeel. Naar aanleiding hiervan is verondersteld dat het onderscheid

tussen stemhebbende en stemloze ploffers primair berust op het waarnemen van de (on)gelijktijdigheid, of (a)synchronie, in de inzet van geluiden met lage (stembandtrilling) hoge (explosieruis) frequenties. Merk echter op dat negatieve steminzet hetzelfde is als aanwezigheid van stembandtrilling tijdens het zgn. stille interval, en dat positieve steminzet identiek is aan aspiratie. Voor zover ik heb kunnen nagaan, zijn er nooit proeven gedaan om deze verwarring van steminzet met andere cues uit te sluiten.

*Wrijfklanken* verschillen van ploffers vooral daarin dat het stille, formantloze, interval bij wrijfklanken geheel wordt opgevuld met ruis. Dit is in beginsel dezelfde ruis, met dezelfde kleuring, die we bij de stemloze ploffers aan het eind van de stilte aantreffen. Het ruis bij wrijfklanken duurt alleen veel langer en de inzet ervan is heel voorzichtig, is de /bruusk/ dan horen we toch weer een ploffer (die daarna overgaat in een wrijfklank). [27] Formantveranderingen voor en na de ruis zijn bij wrijfklanken onbeduidend [27]. Evenals bij ploffers horen we het onderscheid tussen stemhebbende en stemloze wrijfklanken niet zo zeer aan de aan- of afwezigheid van toonhoogte, maar eerder aan verschillen in duur en intensiteit van de ruis.

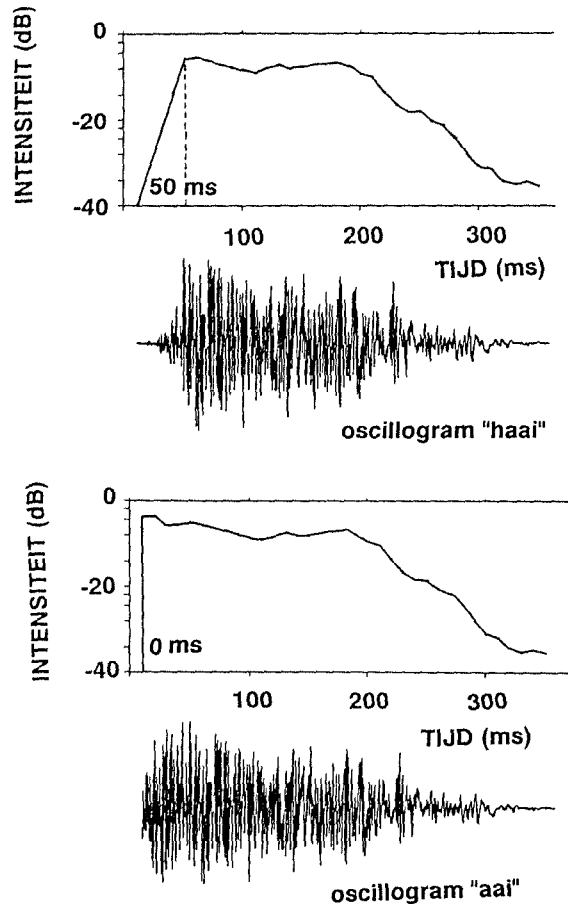
### Sonoranten

Bij *nasalen* wordt tijdens de afsluiting in het mond-keelkanaal de doorgang naar de neusholte open gehouden, waardoor in het gebied tot ca. 1000 Hz nog vrij veel energie hoorbaar blijft. Het neusgeluid zelf is op het gehoor niet of nauwelijks verschillend van bijvoorbeeld [m], [n] en [ŋ]. Het verschil tussen deze drie in articulatieplan wordt vooral aangegeven door de snelle veranderingen in de  $F_2$  voor en na de afsluiting, op de manier waarop we dat eerder zagen bij de plofklanken (zie Figuur 9, onderste paneel).

De zgn. *halfklinkers* [w] en [j] hebben dezelfde formantstructuur als resp. de klinkers [u] en [i], maar de halfklinkervarianten hebben een geringere intensiteit en worden tevens gekenmerkt door heel duidelijke, relatief trage formantveranderingen voor en kort na de (onvolmaakte) afsluiting. In kunstmatige spraak kunnen we een [i] of een [w], of een [d] in een [j] veranderen door alleen de snelheid van de  $F_2$ -verandering te verminderen [16].

De *liquidae* [l] en [r] lijken sterk op elkaar. Zij hebben beide een duidelijke formantstructuur. Hun energie, vooral in de hogere formanten, is geringer dan bij de klinkers, en er zijn duidelijke formantveranderingen op de overgangen van aangrenzende klinkers en de medeklinker. Bij de [l] lopen de  $F_2$  en de  $F_3$  tijdens (gedeeltelijke) afsluiting vrij sterk uiteen, om bij de overgang naar de volgende klinker weer naar elkaar toe te bewegen, de [r] herkennen we vooral aan een verlaging van de  $F_2$ .

De [h] ten slotte is een geval apart. Hij lijkt als twee druppels water op de klinkers, waaraan hij voorafgaat, met precies dezelfde formantstructuur, zij het dat de intensiteit van de hele linie zwakker is. De [h] wordt gemaakt met veel ruis, die wordt opgewekt in de strottehoofd. In flusterspraak is het spectraal onderscheid tussen de [h] en een klinker niet goed als verdwenen. Toch kunnen we nog altijd goed horen of iemand aan of haai flust. Het verschil dat we dan nog over hebben zit hem in de abruptheid waarmee de klinker inzet. Begint het woord met een klinker dan zet die abrupt in, zit er een [h] voor, dan zet de klinker geleidelijk in (zie Figuur 12).



figuur 12 Oscillogram (golfvormregistratie) en intensiteitscurve voor het woordpaar **haai** en **aai**. Het enige verschil wordt gevormd door de inzet van de klinker: een geleidelijke inzet (hier 50 ms stijgtijd) wordt als [h] waargenomen, een abrupte (hier 0 ms stijgtijd) klinkt als een glottisslag (vaste klinkerinzet). Aan deze registraties ligt alleen het gefluisterde woord ten grondslag.

#### DE ROL VAN DE GESPROKEN CONTEXT

Als gezegd zullen we het in dit hoofdstuk niet hebben over woordherkenning. Maar ook als we geen woorden kunnen herkennen in de geluidsstroom die ons oor treft, maken we bij

de identificatie van klanken uitvoerig gebruik van informatie uit de gesproken context. In stel drie soorten contextinformatie aan de orde

## 5.1 SPECTRALE NORMERING

We hebben eerder al gezien dat als twee sprekers voor het gehoor twee precies dezelfde klinkers (of medeklinkers) maken, deze akoestisch flink kunnen verschillen. Over het algemeen is het moeilijk om van een enkele, los gesproken klinker te zeggen welke bedoeld wordt. Identificatie is succesvoller als we een klank laten horen onmiddellijk na een kenniszinnetje waarin de luisteraar zich een indruk kan vormen van de eigenaardigheden van de spreker in kwestie. Hoe kort zo'n kennismakingszinnetje mag zijn, en welke klanken daarin bij uitstek moeten voorkomen, is niet echt bekend. Het vermoeden is dat een zin van slechts enkele woorden al volstaat [9].

## 5.2 TEMPORELE NORMERING

Als mensen snel spreken, dan gebeuen er verschillende dingen. Door de bank genomen zal de duur van elke aparte spraakklank geringer worden, en tevens zullen de spraakklanken minder vol, dus onnauwkeuriger worden uitgesproken. Dit betekent dat de criteria die we gebruiken om bijvoorbeeld te beslissen of we met een korte of een lange klinker te doen hebben, niet vastliggen, maar verschuiven naar gelang het spreektempo. Stel dat bij een normaal spreektempo de grens tussen korte en lange klinkers ligt bij 100 ms. Klinkers die korter duren dan 100 ms worden dan als korte klinkers geïdentificeerd, en klinkers boven dit criterium als lange. Bij een hoger spreektempo moet dit criterium bij een lagere waarde komen te liggen, bijvoorbeeld bij 90 of zelfs bij 80 ms. Voordat een luisteraar de duur van een klinker kan gebruiken bij zijn beslissing over de identiteit van die klinker, moet hij eerst een betrouwbare indruk hebben van het spreektempo van het omringende stuk spraak. Hierbij blijken vooral de klanken in de onmiddellijke nabijheid van belang, en is merkwaardig genoeg - het tempo van de volgende klanken van grotere invloed dan dat van de voorafgaande klanken. Er is dus sprake van achterwaartse temporele normering. Dit betekent opnieuw dat beslissingen over de identiteit van individuele klanken in de taal alleen achteraf genomen kunnen worden [33].

## 5.3 COARTICULATIE

Bij het spreken verspringen onze spraakorganen niet van de ene klank naar de volgende, maar veranderen ze min of meer vloeiend van de ene doelpositie naar de andere, zonder ooit het beoogde doel te bereiken. Als gevolg hiervan kunnen we aan de spectrale verandering meestal horen welke klank bezig is gesproken te worden, welke klank er volgt, en welke klank gaat volgen. Een spraakklank bevat daarmee niet alleen informatie over zijn eigen identiteit, maar ook over die van zijn burens. De aanwezigheid in een spraakklank van eigenschappen van aangrenzende klanken noemen we *coarticulatie*. De onderlinge beïnvloeding tussen spraakklanken zal sterker zijn naarmate de betrokken klanken in de taal dichterbij elkaar staan. Kunstmatige spraak waarin nagelaten is de effecten van

coarticulatie tussen naastliggende klanken te simuleren, is in het algemeen slecht te verstaan

Er is al veel onderzoek gedaan naar de rol van coarticulatie op de herkenning van spraakklanken. Wanneer we de begin- en eindmedeklinker in een CVC-syllabe via een elektronische ingreep onhoorbaar maken, dan kunnen luisteraars de weggevalen medeklinkers toch identificeren, voornamelijk aan de hand van de snelheid en richting van formanttransities (zie ook boven). Hierbij geven de *eindtransities* grosso modo nog wat meer houvast dan de *begintransities*. Omgekeerd kunnen we in de ruisstoot van een plof- of wrijfklank goed horen wat voor klinker daar onmiddellijk aan grenst. Voor literatuuroverzichten zie [18, 27].

Los uitgesproken klinkers blijken minder goed geïdentificeerd te worden dan dezelfde klinkers gesproken in CVC-syllaben, ook als we de luisteraar er niet bij vertellen welke medeklinkers het zijn [41]. Het lijkt er dus op dat de klankovergangen bij de spraakwaarneming bruikbaarere informatie verschaffen dan de middens van klanken.

De onderlinge beïnvloeding tussen niet-aangrenzende klanken, als die al meetbaar aanwezig is, heeft alleen in uitzonderlijke situaties nog een merkbaar effect op de spraakwaarneming [18].

## 6 IS ER IETS BIJZONDERS MET SPRAAK?

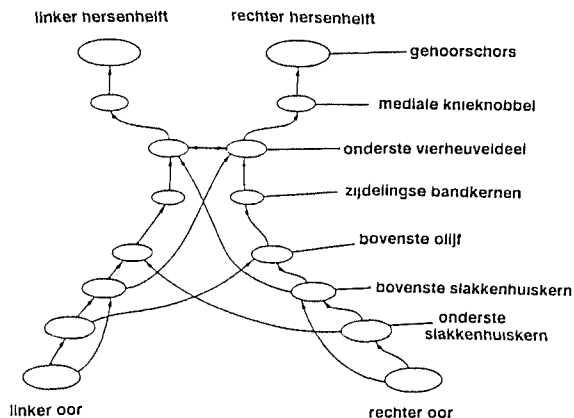
In de afgelopen 20-30 jaar lezen we voortdurend claims die inhouden dat er iets bijzonders aan de hand is met de waarneming van spraak. Spraak zoudeft we op een fundamenteel andere wijze waarnemen dan niet-spraakgeluiden. Ik verdeel deze wereld onder in drie ampden: de *gelovigen*, de *ongelovigen*, en de *agnosten*. Ik reken mezelf voorlopig tot de laatste categorie. Laten we maar eens kijken wat er aan gegevens beschikbaar is.

### 1 RECHTEROORVOORDEEL

we hebben in sectie 2 gezien dat elke haarcel langs het basilaire membraan gevoelig is voor een specifieke frequentie. Elk van deze ca. 30.000 frequentiedetectoren is via de centrale hoorzenuw verbonden met de hersenen. Deze verbinding is niet rechtstreeks, maar loopt via talrijke tussenstations (zie Figuur 13).

We zullen ons niet wagen aan een bespreking van de diverse tussenstations en hun lang voor het spraakverstaan (waarvan nog maar heel weinig bekend is). Belangrijk is dat we constateren dat het systeem in duplo is uitgevoerd: dat is ook redelijk, want we hebben eindelijk twee oren. Onze hersenen zijn verdeeld in een linker- en een rechterhelft (de *hersenhalften* of *hemisferen*, zie ook KOLK). Merkwaardig genoeg heeft de linkerhelft in ons lichaam (spieren en zintuigen) het sterkst contact met de rechter hersenhelft, en respondeert de rechter helft van het lichaam vooral met de linker hersenhelft. De koppeling van lichaamshelft en hersenhelft is dus gekruist. Zo ook voor het gehoor. Slechts een klein deel van de informatie van het linkeroor bereikt de linker hersenhelft, het grootste ergens op de weg naar boven over naar het parallelle kanaal (zie Figuur 13).





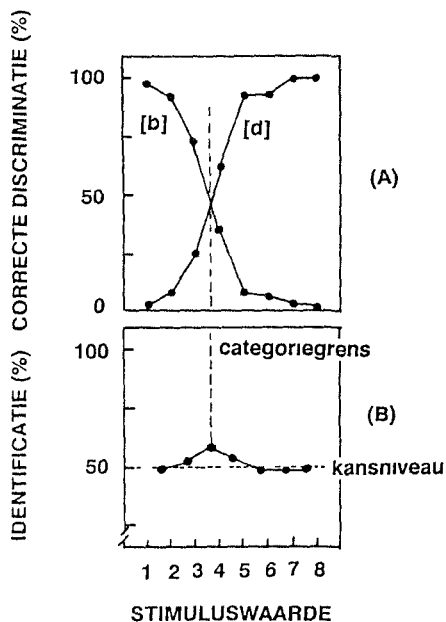
Figuur 13 De afzonderlijke schakelstations in de gehoorbaan (de weg van oor naar hersenschors)

Er zijn aanwijzingen dat de linker hersenhelft gespecialiseerd is voor de verwerking van taal en spraak. Bekend is dat mensen met beschadigingen aan de linker hersenhelft eerder aan taal- en spraakstoornissen leiden dan patienten met een beschadiging aan de rechter hersenhelft. Dit doet vermoeden dat spraakgeluid efficiënter verwerkt zal worden wanneer het binnenkomt via het rechteroor. Inderdaad blijken spraakgeluiden sneller en beter herkenkend te worden als ze alleen met het rechteroor gehoord worden dan alleen met het linkeroor. Met andere geluiden, bijvoorbeeld muziek, ligt dit juist andersom, vandaar dat mensen het rechter- en linkeroor wel aanduidt als resp. het spraakoor en het muziekoor.

Als nu de verwerking van spraak plaatsvindt in een gespecialiseerd deel van het brein, dan ligt het voor de hand te veronderstellen dat bij de verwerking van spraakgeluid een bijzonder mechanisme betrokken is. Toch hoeft dit niet per se zo te zijn. Het is nog altijd mogelijk dat de verwerking van het spraakgeluid net zo verloopt als van elk ander geluid, in dat het taalcentrum in de linker hersenhelft pas ingeschakeld wordt op het moment dat de luisteraar een taalgebonden beslissing moet nemen over het gehoorde, bijvoorbeeld of hij aan de proefleider moet melden wat voor een klinker of medeklinker hij meende te herkennen.

## 2.2 CATEGORIALE WAARNEMING

We nemen van een spreker de lettergrepen [ba] en [da] op band op. Vervolgens maken we via een elektronische ingreep een reeks nieuwe lettergrepen die het verschil tussen [ba] en [da] (het zgn. continuum) overbruggen in bijvoorbeeld 7 kleine, gelijke stappen. We hebben dan 8 lettergrepen, genummerd 1 (de oorspronkelijke [ba]) t/m 8 (de oorspronkelijke [da]), waarvan elk hoger nummer een klein beetje minder op [ba] lijkt en een beetje meer op [da]. We laten deze lettergrepen in willekeurige volgorde horen aan luisteraars die steeds moeten beslissen of ze [ba] of [da] horen. De resultaten van zo'n roet zijn te zien in Figuur 14.



Figuur 14 A categorisatie en B discriminatie langs een [ba] - [da] continuum. Categoriële waarneming is duidelijk alleen als de syllaben in een vergelijkingspaar uit verschillende tonetische categorieën komen, worden ze (boven kans) gediscrimineerd (naar Studdert-Kennedy, 1976)

Bijna alle lettergrepen, zowel de oorspronkelijke als de elektronisch bijgemaakte, delen de luisteraars eensgezind in als hetzij [ba] hetzij [da]. De stemmen staken alleen voor de nummers 3 en 4, precies midden tussen de uitersten in. Luisteraars zijn kennelijk voortreffelijk in staat de verscheidenheid aan geluiden te sorteren in twee duidelijk onderscheiden categorieën [ba] en [da].

Vervolgens maken we paren van steeds twee minimaal verschillende lettergrepen, dus niet onmiddellijk opeenvolgende nummers [1, 2], [2, 3], [7, 8]. De luisteraars moeten nu eggen of ze wel/geen verschil kunnen horen tussen de lettergrepen van elk paar. Het blijkt zie ook Figuur 14) dat luisteraars alleen goed kunnen horen dat die lettergrepen hoorbaar aan elkaar verschillen als die eerder waren ingedeeld in verschillende categorieën [ba - da], waren lettergrepen die eerder waren toegewezen aan dezelfde categorie, klinken identiek. Dit verschijnsel nu wordt *categoriële waarneming* genoemd.

De suggestie is nu dat zulke categoriële waarneming alleen voorbehouden is aan praakklanken, met name plofklanken. Daartegenover moeten we bedenken dat juist plofklanken zwakke, kortdurende en snel-veranderende geluiden zijn, die snel uit het uithoofd geheugen vervagen. Om de informatie uit zulke geluiden te kunnen vasthouden moet de luisteraar de plofklank provisorisch indelen in een of andere categorie, waarbij de

ekende fonetische categorieën een uitstekend aanknopingspunt bieden. Deze situatie zou zich ook kunnen voordoen bij andere geluiden dan spraakklanken.

We zouden dus een proef moeten uitvoeren met niet-spraakachtige geluiden die eveneens de eigenschap hebben snel uit het auditief geheugen te vervliegen, en die gedeeld kunnen worden in ons vertrouwde, herkenbare categorieën. Hierbij valt te denken aan het geluid van hamerslagen op hout, steen of metaal. Zo'n proef wordt inmiddels uitgevoerd [38], maar resultaten zijn nog niet beschikbaar. Overigens is al wel categoriële waarneming gevonden voor de herkenning van muzikale accoorden als A-majeur of A mineur [4, p. 216], zodat de houdbaarheid van categoriële waarneming als unieke eigenschap van spraakperceptie bij voorbaat gering is.

### 3 CUE TRADING

We hebben boven al opgemerkt dat een klankcontrast zelden teruggevoerd kan worden op slechts een akoestisch verschil. Er zijn in de regel meerdere eigenschappen van het spraakgeluid die systematisch verschillen tussen de contrasterende klanken (sectie 4.3). Zou men we dat de kleur en duur van een klinkergeluid in het Nederlands beide van belang zijn voor de beslissing of we een korte [α] of een lange [a] horen. Een klinkergeluid dat op grond van zijn duur een lange [a] zou moeten zijn, kunnen we veranderen in een korte [α] door de kleur in de richting van [α] op te schuiven. Dergelijke onderlinge compensatiemogelijkheden tussen de vaak heel diverse akoestische eigenschappen (*cues*) van spraakklanken hebben veel aandacht gekregen in de literatuur, en staan bekend onder de naam *trading*, letterlijk *uitwisseling*. De claim nu is dat *cue trading* alleen voorkomt bij de identificatie van spraakklanken.

Een nog verdergaande claim is dat de vele, uiteenlopende verschijningsvormen van een klank toch herkend worden als steeds dezelfde categorie, doordat de luisteraar zich voortdurend indenkt hoe hij zo'n klank als spreker zou maken. Zo bezien is categoriële waarneming een rechtstreeks gevolg van *categoriële produktie*. We kunnen op het gehoor een klanken onderscheiden tussen [b] en [d] in, omdat we er ook geen klanken tussen kunnen maken: de [b] maken we met de lippen, de [d] met de tongpunt, en iets er tussenin bestaat niet. Volgens deze zgn. *motortheorie van spraakwaarneming* maakt de luisteraar gebruik van het binnenkomend spraakgeluid een reconstructie van het spraakproductieproces aan dat geluid ten grondslag ligt. De luisteraar praat onmerkbaar mee met de spreker. Ook moderne uitgaven van deze theorie gaan ervan uit dat de luisteraar met behulp van alle beschikbare informatie, hoorbaar, zichtbaar, of zelfs voelbaar, een reconstructie maakt van de articulatorische gebeurtenis (*event*) die plaatsvindt. Volgens aanhangers van deze theorie, de zgn. *event perception theory*, verloopt zo'n reconstructie onbewust, onwillekeurig, onmiddellijk, en altijd trefzeker [20].

Tegen de gedachte van de motortheorie is vaak ingebracht dat ook mensen die vanaf de geboorte stom zijn, maar niet doof, een volstrekt normale spraakwaarneming vertonen inclusief categoriële perceptie en *cue trading*. Deze mensen kunnen echter nooit een reconstructie maken van de articulatorische gebeurtenis. Los daarvan lijkt het mij heel goed mogelijk *cue trading* aan te tonen bij de categorisatie van niet-spraakklanken. Alweer: dat dit type proeven wordt gewerkt, maar er zijn nog geen resultaten beschikbaar.

## 6.4 AANGEBOREN SPRAAKKENMERKDETECTOREN

De gehoorcellen detecteren de aanwezigheid van specifieke frequenties in het geluid. Deze informatie wordt vanaf het gehoor naar de hersenen doorgegeven via een keten van tussengeschakelde zenuwcellen (zie Figuur 13). De keten van cellen is hiërarchisch georganiseerd, en wel zo dat een hoger gelegen cel informatie uit twee of meer lagere cellen combineert, en aldus kan fungeren als een gespecialiseerde patroonherkenner. Zo is verondersteld dat zich hoog in de hiërarchie gespecialiseerde cellen bevinden die elk in staat zijn de aanwezigheid van een specifiek klankkenmerk te detecteren. Zo zouden er detectoren zijn voor de kenmerken stemhebbend versus stemloos, en voor de plaats van articulatie (labiaal, alveolair, velair). Men vermoedt het bestaan van zulke kenmerkdetectoren op grond van het verschijnsel *gewenning* (of *adaptatie*). Hiermee wordt ruwweg het volgende bedoeld:

Stel weer dat onder normale omstandigheden de scheidslijn tussen een korte en een lange klinker ligt bij 100 ms (zie 5.2). Als we nu een luisteraar gedurende langere tijd steeds overduidelijke voorbeelden geven van ("bombarderen met") extreem korte klinkers, dan zal blijken dat hij daarna een wat langere klinker benoemt als een lange klinker, ook al is deze in werkelijkheid korter dan 100 ms. We noemen dit *criteriumverschuiving* als gevolg van *gewenning*. De gedachte is nu dat de kenmerkdetector door het bombardement van extreem korte klinkers ongevoelig is geworden voor korte klinkers, waardoor de luisteraar eerder zal beslissen dat een klinker lang is.

Het is gemakkelijk aan te tonen dat zo'n kenmerkdetector niet in het oor zelf zit, als we het bombardement met extreem korte voorbeelden via het ene oor laten horen en de langere klinker via het andere, vinden we dezelfde criteriumverschuiving. De kenmerkdetector zit dus ergens waar de informatie uit beide oren al samengekomen kan zijn.

Ook al zouden er speciale detectorcellen bestaan, afgestemd op eigenschappen van spraakgeluiden, dan nog maakt dat spraak niet echt bijzonder. Op dezelfde manier zouden we ook kunnen beschikken over detectoren voor bepaalde muziekinstrumenten, voor dierengeluiden en andere veel-voorkomende geluiden uit onze dagelijkse omgeving. Daarom heeft het debat zich gaandeweg toegespitst op de vraag of de spraakdetectoren aangeboren zijn, d.w.z. vanaf de geboorte aanwezig. Inderdaad vertonen ook pasgeboren kinderen verschijnselen van *gewenning* en categoriale waarneming bij spraakgeluiden (KOOPMANS-VAN BUNUM & VAN DER STELT). Sterker nog, dezelfde verschijnselen zijn ook gevonden bij chinchillas, hetgeen erop duidt dat zoogdieren in het algemeen een choormechanisme bezitten dat gespecialiseerde detectoren heeft voor bepaalde geluiden. Dit op zijn beurt leidt tot de conclusie dat niet ons gehoor zich heeft aangepast aan de eisen die spraakverstaan stelt, maar omgekeerd, dat we in de loop van de evolutie (spraak)geluiden zijn gaan gebruiken waarvoor ons zoogdierengehoor toch al een bijzondere gevoeligheid heeft.



## 5 SLOT

De gedachte dat de waarneming van spraak gebruik maakt van een speciaal mechanisme is op het eerste gezicht aantrekkelijk, en lijkt veel merkwaardige verschijnselen begrijpelijk te maken. Daartegenover geldt in de wetenschap de ijzeren wet van Occam's scheermes: al een deel van een verklaring kunt weglaten en toch dezelfde verschijnselen kunt verklaren, dan is die kortere verklaring de beste. Het is niet onmogelijk dat spraak wordt waargenomen met een speciaal mechanisme, bijvoorbeeld via reconstructie van de articulatie of via overgeefde gespecialiseerde kenmerkdetectors. Vooralsnog echter kunnen alle eigenaardigheden van de spraakwaarneming begrepen worden uit algemene eigenschappen van het menselijk gehoor en de organisatie van het geheugen. De toekomst zal leren wie er uiteindelijk gelijk heeft.

## ITERATUUR

### VERZICHTSWERKEN

- [1] Ainsworth W.A. (1976) *Mechanisms of speech recognition*. Oxford: Pergamon Press.
- [2] Darwin C.J. (1976) The perception of speech in Cutcliffe L.C. & Friedman M.P. (eds.) *Handbook of perception volume VII: Language and speech*. New York: Academic Press, pp. 175-226.
- [3] Green D.M. (1976) *An introduction to hearing*. Hillsdale, NJ: Erlbaum. voorl. hoofdstuk XII (Speech perception).
- [4] Moore B.C.J. (1982) *An introduction to the psychology of hearing*. London: Harcourt Brace & Jovanov. hoofdstukken I t/m IV en VII (Speech perception).
- [5] Nootboom S.G. & Cohen A. (1984) *Spreken en verstaan: een nieuwe inleiding tot de experimentele fonetiek*. Assen: van Gorcum. hoofdstuk III 5 (Spectrografische afbeelding van spraakgeluiden) III 6 (De proef of synthese van spraak) IV (Het oor en het horen van geluiden).
- [6] Schouten M.E.H. (ed.) (1987) *The psychophysics of speech perception*. The Hague: Nijhoff.
- [7] Studdert-Kennedy M. (1976) Speech perception in Lass N.J. (ed.) *Contemporary issues in experimental phonetics*. New York: Academic Press, pp. 243-293.

### WETENSCHAPPELIJKE REFERENTIES

- [1] Abel S.M. (1972) Discrimination of temporal gaps. *Journal of the Acoustical Society of America* 52: 519-5.
- [2] Balen C.W. van (1980) *Intelligibility of speech fragments*. Dissertatie, RU Utrecht.
- [3] Benjumeil A.P. & D'Arcy J. (1986) Time warping and the perception of rhythm in speech. *Journal of Phonetics* 14: 231-246.
- [4] Bladon R.A.W. & Lindblom B.L. (1981) Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America* 69: 1414-1422.
- [5] Bot C.L.J. de (1982) *Visuele feedback van intonatie*. Dissertatie, KU Nijmegen.
- [6] Breuwer M. (1985) *Speech reading, supplemented with auditory information*. Dissertatie, VU Amsterdam.
- [7] Broecker M.P.R. van den & Heuven V.J. van (1983) Effect and artifact in the auditory discrimination of onset and decay time: speech and non-speech. *Perception and Psychophysics* 33: 305-313.
- [8] Brox J.P. & Nootboom S.G. (1982) Intonation and the perceptual separation of simultaneous vowels. *Journal of Phonetics* 10: 23-36.
- [9] Ducloux R. (1976) Literature analyzers for the phonetic dimension: stop vs. continuant. *Perception and Psychophysics* 19: 267-272.
- [10] Dinnsen S.J. (1980) Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America* 67: 253-261.
- [11] Dupuis M.Ch. (1987) *Perceptual effects of phonetic and phonological accommodation: an experimental study on effects of coarticulation and assimilation on perception of words and word beginnings*. Dissertatie, RU Leiden.
- [12] Fanning J.L. (1955) A difference limen for vowel formant frequency. *Journal of the Acoustical Society of America* 27: 613-617.
- [13] Fowler C.A. (1986) An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics* 13: 3-28.

- [21] Fujisaki H, Nakamura K & Imoto T (1975) Auditory perception of duration of speech and non speech stimuli in Fant G & Tatsumi M A A (eds) *Auditory analysis and perception of speech* New York etc Academic Press pp 197-219
- [22] Gerstman L J (1957) *Perceptual dimensions for the friction portion of certain speech sounds* Dissertation New York University
- [23] Hart J C (1981) Differential sensitivity to pitch distance particularly in speech *Journal of the Acoustical Society of America* 69 811-821
- [24] Hawkins S & Stevens K N (1985) Acoustic and perceptual correlates of the non nasal distinction for vowels *Journal of the Acoustical Society of America* 77 1560-1575
- [25] Heuven V J van & Broekke M P R van den (1979) Auditory discrimination of rise and decay times in tone and noise bursts *Journal of the Acoustical Society of America* 66 1308-1315
- [26] Heuven V J van Houten J E van & Vries J W de (1986) De perceptie van Nederlandse klinkers door Turken *Spektator* 15 225-238
- [27] Klaassen Don L L O (1983) *The influence of vowels on the perception of consonants* Dissertation RU Leiden
- [28] Liberman I (1970) *Suprasegmentals* Cambridge MA MIT Press
- [29] Lindblom B E F (1980) Phonetic universals in vowel systems in Ohala J J (ed) *Experimental phonology* New York etc Academic Press pp 13-44
- [30] Massaro D (1988) Speech perception by ear and eye in Dodd B & Campbell R (eds) *Hearing by eye* *Experimental studies in the psychology of lipreading* Hillsdale NJ Erlbaum
- [31] Mennelstein P (1978) Difference limens for formant frequencies of steady state and consonant bound vowels *Journal of the Acoustical Society of America* 63 572-580
- [32] Nootboom S G (1974) Experimentele bijdragen aan de fonologie *Forum der Letteren* 15 73-99
- [33] Nootboom S G (1981) Speech rate and segmental perception or the role of words in phonemic identification in Myers F, Laver J & Anderson J (eds) *The cognitive representation of speech* Amsterdam North Holland pp 143-150
- [34] Nord L & Svenclius C (1977) Analysis and prediction of difference limen data for formant frequencies *Quarterly Progress and Status Report Speech Transmission Laboratory Royal Institute of Technology Stockholm* 3/4 pp 60-72
- [35] Pols L C W & Schouten M E H (1987) Perception of tone band and formant sweeps in [6] pp 231-240
- [36] Remez R E, Rubin P E, Pisoni D B & Carrell J D (1981) Speech perception without traditional speech cues *Science* 212 947-950
- [37] Rietveld A C M & Gussenhoven C (1985) On the relation between pitch excursion size and prominence *Journal of Phonetics* 13 299-308
- [38] Schouten M E H (1987) Speech perception and the role of long term memory in [6] pp 66-79
- [39] Shis I H & Nicrop D J P J van (1970) On the forward masking threshold of vowels in VC combinations *IPO Annual Progress Report* 5 68-77
- [40] Shis I H & Cohen A (1969) On the complex regulating the voiced voiceless distinction I & II *Language and Speech* 12 80-102 137-155
- [41] Strange W, Verbrugge R R, Shankweiler D P & Edman T R (1976) Consonant environment specific vowel identity *Journal of the Acoustical Society of America* 60 213-224
- [42] Verbrugge R R, Strange W, Shankweiler D P & Edman T R (1976) What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America* 60 198-212
- [43] Resnick S B, Weiss M S & Heinz J M (1977) Masking of filtered noise bursts by synthetic vowels *Journal of the Acoustical Society of America* 66 674-677

## /RAGEN

Wat is de functie van het middenoor? Hoe zou het komen dat we bij een snelle en forse hoogteverandering (zoals in een dalend of opstijgend vliegtuig) tijdelijk hordhoïend zijn?

Wat is het gevolg van een frequentieverandering voor het bewegingspatroon van het basilaire membraan? En wat van een intensiteitsverandering?

Hoe noemen we het frequentiegebied waarbinnen gelijktijdige tonen elkaars waarneming beïnvloeden? Hoe groot is dat gebied ruwweg?

Wat is een Juist Waarneembaar Verschil? Hoe groot is het JWV (in absolute getallen of in een percentage) voor intensiteit, grondfrequentie van een klinker, formantfrequentie van een klinker en duur van een klinker?

Wat zou het effect zijn van gelijktijdige (zijwaarts) maskering op de waargenomen scherpte van klinkerformaten?

Als we het Nederlandse woord *aak* op een bandje opnemen en dit achterstevoren afspelen, dan horen we niet *kaak* maar gaat er eventueel chaos zoals in het woord *chaos*. Kunt u deze verandering van plofklank in vrijklank verklaren?

De echte medeklinkers (de *zgn* obstruenten) kunnen worden onderverdeeld in een groep stemhebbende en een groep stemloze medeklinkers. Zo staan tegenover elkaar de klanken [p b], [t d], [k g] en [s z]. Hoe kunt u we gemakkelijk aantonen dat dit onderscheid niet in de eerste plaats berust op de aanwezigheid of afwezigheid van stembandtrilling? Welke andere verschillen tussen de beide klankcategorieën zijn belangrijker?

Waarom horen we dat een spraakklank een klinker is? Is het horen van stemhebbendheid hierbij belangrijk? Waarom wel niet?



9. Wat wordt bedoeld met cue-trading in de spraakwaarneming? Kunt u voorbeelden geven van cue-trading?
10. Hoe horen we bij plofklanken wat de articulatieplaats is?
11. Wat bedoelen we met coarticulatie?
12. Los-gesproken klinkers worden voller en duidelijker van elkaar verschillend uitgesproken dan dezelfde klinkers in CVC-syllaben. Toch zijn ze in een syllabecontext beter te herkennen. Hoe zou dat kunnen komen?
13. Hoe kunnen we aantonen dat de luisteraar bij het spraakverstaan ook gebruik maakt van visuele informatie van de spreker hem verschaft?
14. Hoe kunnen we gemakkelijk laten zien dat kenmerkdetectors voor bijvoorbeeld stemhebbendheid, als die bestaan, niet zetelen in het oor zelf, maar meer centraal in de hersenen?
15. Als we in Figuur 5 (bovenste paneel) de harmonischen met even nummers zouden weglaten, welke toonhoogte zou ons gehoor dan toekennen aan het resterende tooncomplex? En wat zou er gebeuren als we juist de oneven harmonischen weglaten?